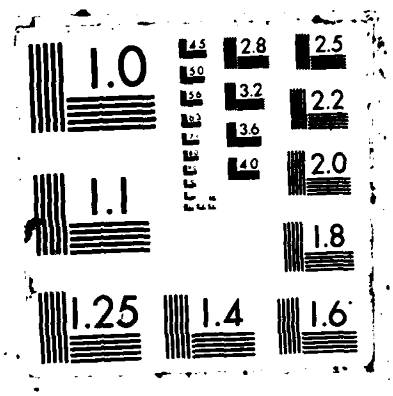
AD-A183 460 PRODUCTIVITY: EFFECTS OF INFORMATION FEEDBACK ON HUMAN ERRORS(U) INSTITUTE FOR BEHAVIORAL RESEARCH INC SILVER SPRING MD H M PARSONS 31 JUL 80 N00014-77-C-0024 1/1

UNCLASSIFIED F/G 5/2 NL

END
9-87
DTIC


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

SEP 11 1980

(1)

PRODUCTIVITY: EFFECTS OF INFORMATION FEEDBACK ON HUMAN ERRORS

H. McIlvaine Parsons

FINAL REPORT

July, 1980

Preparation of this report was supported by the Engineering Psychology Program, Office of Naval Research

ONR Contract N00014-77-C-0024 Work Unit NR197-044

87

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A183 460

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Final Report	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Productivity: Effects of Information Feedback on Human Errors		5. TYPE OF REPORT & PERIOD COVERED Final Report 1 Oct. 1977 - 31 Jan. 1979
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) H. McIlvaine Parsons		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0024
9. PERFORMING ORGANIZATION NAME AND ADDRESS Institute for Behavioral Research 2429 Linden Lane, Silver Spring, MD 20910		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 197-044
11. CONTROLLING OFFICE NAME AND ADDRESS Engineering Psychology Program Office of Naval Research - Code 455 Arlington, VA 22217		12. REPORT DATE July 31, 1980
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

	Copying
Information Feedback	Detection
Human Errors	Coding
Productivity	Reinforcement
Information Processing	Consequation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes a research program exploring the effects of verbal consequators (information feedback or reinforcers/punishers) on human errors in copying, detection, and coding tasks. The program undertook three experimental studies, a study of consequator ratings, and a mapping of consequation variables, but failed to produce evidence of differential effects, though much was learned about the methodology of inducing errors experimentally. The ratings study has implications for job performance.

DD FORM 1 Jan 73 1473

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TABLE OF CONTENTS

INTRODUCTION

In the growing concern about productivity in both military and civilian domains, error reduction has been overshadowed by interest in output and attendance, in productivity research. The project described in this final report on Contract N00014-78-C-0024 investigated error reduction in productivity-related human information processing tasks. Such tasks can be critical to all kinds of naval operations and maintenance. The tasks initially selected for examination in this project were copying, detecting, coding, and categorizing; the last was dropped. Considerable experimental effort was devoted to the others, though it must be viewed as exploratory rather than definitive. Outside the psychomotor and sensory fields, relatively little experimental research has been directed explicitly at variables influencing human error in repetitive information processing performance (rather than learning). Such is especially the case for within-subject designs. Whenever a new domain is studied, more exploration may be needed than was anticipated, as in this project.

The category of independent variables of interest can be labelled as "consequation," which may be interpreted to include both information feedback and reinforcement. Both of these terms mean that an event that follows some performance may influence its future occurrence. A very large number of variables are involved in the effects of consequation, as a technical report published as part of this project attests (Parsons, 1979). The form of consequation examined in the project was verbal, in particular verbal feedback that might be viewed as either informational, or derogatory/complimentary, or both. Remarkably little investigation has been accorded aversive verbal consequation (other than right-wrong studies of knowledge of results), though unfavorable comments are common events in organizational, family, and other situations.

From a theoretical viewpoint, the project reflected an interest in relating to each other the cognitive area of information feedback (or knowledge of results) and the behavioral area of reinforcement (and punishment) in operant conditioning. The information an individual receives after and about performance and some hedonic consequence, reinforcing or punishing, may occur together either from separate circumstances or in the same verbal comment from another person. Some cognition-oriented analysts (e.g., Annett, 1969) have preferred to attribute all effects of consequation to its informational content. Behaviorists tend to give the entire credit to motivational variables, reinforcers or punishers, even though their views of behavior include discriminative stimuli as behavioral outcomes (and discriminative stimuli, or "discriminators," can be considered "information"). In reviews of the Hawthorne studies (Parsons, 1974, 1978), the principal investigator in this project concluded that the Hawthorne effect resulted from a combination of information feedback (discriminator consequation) and monetary reward (reinforcement consequation) contingent on performance that increased productivity. His Hawthorne analyses were responsible for proposing the present project, though it was directed at error reduction rather than an increase in output.

Both informational and motivational variables were included in the technical report already mentioned and described briefly in this final report as Mapping of Consequation Variables, and they are implied in the section on Consequator Rating Study, an inquiry into self-reports about such variables. Relationships between the two types of variables constituted the underlying reason for the three experimental studies undertaken and reported, though the two studies which actually manipulated consequation as a variable must be regarded as only a preliminary and exploratory step toward examining such relationships. The studies are described under the headings Copying, Detection, and Coding, the three information processing tasks mentioned earlier.

The copying study failed to produce evidence that information feedback reduced the frequency of errors or that differences in the nature of the feedback affected error frequency. Following a substantial exploratory phase, data were obtained from only two subjects in the detection study before this report was submitted, and these data had not been processed. The coding study also had many exploratory sessions to explore methods of inducing error but never reached the point of examining different types of information feedback. Thus, much was learned about methodology for inducing errors in human information processing but the project failed to demonstrate differential effects from information feedback.

Among the lessons learned was the difficulty of inducing error at levels that would make it possible, without an exorbitant number of trials, to determine whether differences in results from experimental conditions were consequential and significant. Due probably to the demand characteristics of the experimental arrangements, subjects tended to try to keep from making errors much as they had done in tests at school. As a result, the tasks had to be made more difficult or the trials and sessions longer or more numerous than otherwise might have been necessary, especially in a within-subjects design. Though more subjects would be needed and the costs would be greater, a between-subjects design may be viewed as preferable for this sort of investigation.

During the course of this project but not as an integral part of it, the principal investigator gave an invited paper at the Symposium on Productivity Enhancement: Personnel Performance Assessment in Navy Systems, of the Navy Personnel Research and Development Center, San Diego (Parsons, 1977).

Appreciation is extended to the project's Scientific Officer, Dr. Martin A. Tolcott, Director, Engineering Psychology Program, for his patience with unavoidable problems that impeded the project's timely completion, and to Mr. Kenneth R. Williams, Research Assistant at the Institute for Behavioral Research, whose collaboration in collecting data and in other aspects of the project was conscientious and invaluable.

All participants in the research signed informed consent forms, which, together with the procedures, were approved by the Committee for the Protection of Humans in Research, the institutional review board of the Institute for Behavioral Research.

COPYING STUDY

A familiar human information processing task is copying or transcribing, either between input modalities or within a modality. In the latter instance an individual may read from one kind of visual presentation and copy the material, without recording it, on another kind of visual display. It is common experience that errors occur in copying. A study was conducted to investigate the effects of several kinds of consequation, or information feedback, on copying errors.

The experimental strategy was to use a within-subjects design and have a small number of participants (subjects). No large pool of relatively homogeneous individuals was readily available for selecting randomly a substantial number of participants for different treatment conditions. Since with a within-subjects design it is essential to minimize learning that could confound the effects of the states of the independent variable(s), it seemed necessary to select material to copy that already was well learned, that is, was well established in the subject's repertoire. It was also advisable to schedule a large number of trials prior to the data-taking trials.

Since most Americans who have at least graduated from high school have extensively discriminated and manipulated numbers, learning of individual numerals will not occur during an experiment. (The participants in this study responded to advertisements posted at a nearby junior college, although not all were students.) Numbers can be selected and arranged sequentially by means of random number tables so that sequences and arrays will be different but equivalent. Their differences should obviate learning in the form of memorization of sequences or arrays in repeated presentations during the study, and their equivalence should prevent confounding due to differences in the stimulus material. For these reasons rows of random numbers consituted the material to be copied.

Participants

Ten participants (college students or their acquaintances) performed in the experiment proper, six men and four women. They were told that the study involved clerical tasks but were not informed about the information feedback conditions. They signed consent forms approved by the Committee for Protection of Humans in Research (institutional review board) of the Institute for Behavioral Research, which approved the procedures in the study. Seven other participants took part in an earlier exploration of procedures.

Procedures

Each participant performed in a three-hour session, and six of them performed in a similar session six to 11 days earlier. Each of the other four had taken part in a session seven to 25 days earlier when they carried out a similar number copying task in the procedural exploration.

Each participant encountered five conditions. The first and last were baseline conditions, without feedback. In the others, information feedback took three forms: "You got most of them right," "You got some of them wrong," and "You got some of them wrong. Mistakes with numbers can be most unfortunate in some clerical tasks." There were five orders of encountering these conditions among the ten participants.

In each condition, a participant had to copy, on a typewriter, random numbers that were projected on a white surface from a transparency. The numbers were arranged in columns or blocks of ten rows (two sets of five) each. For seven of the participants there were eight numerals per row, for the other three nine, except in the first two baseline blocks, which had ten. Accuracy performance on these two was examined to determine whether a participant should copy eight or nine numerals in a row, nine being more difficult. Each condition had five blocks or columns of numerals with eight or nine per row. Thus, a participant copied 400 or 450 numerals per condition (plus the additional two blocks in the initial baseline). Only one row was displayed at a time, for five seconds. The participant had to wait ten seconds before starting to type (after receiving a "Go"). A participant could pace himself/herself in typing but was given the visual instruction, "Type the numbers as fast as you can," at the start of each experimental condition. Participants were not permitted to make corrections. A new row of numerals was displayed when the participant pressed the carriage return. The participant sat on the other side of a partition from the investigator, facing the projected display. A lamp illuminated the typewriter, and overall illumination was kept constant.

There were two-minute intervals between blocks/columns and five minutes between conditions. In the feedback condition, a feedback message was projected for 30 seconds after each block of numerals, beginning 60 seconds after the start of the interval (to make it appear plausible that the investigator had actually counted the number of errors). The time to complete each block was recorded but the errors were not counted until after the session (except for the two initial baseline blocks). Though the feedback messages were not adjusted to actual performance, "most right" and "some wrong" did match that performance.

Within a session, the rows of random numbers in any block were unique; there was no repetition of rows. Numbers were drawn from a random number table. All participants encountered the same blocks but these were assigned differently among conditions for different subjects. Virtually all of the blocks in the principal session had been used in the prior session, but with a different assignment to conditions for the same participant.

At the end of the principal session, each participant answered eight questions about the study. These included rating the feedback conditions on a 7-point unpleasant-pleasant scale.

Procedural Exploration

For two reasons it appeared essential to have a procedure that would actually result in some errors consistently in each participant. One reason was the content of feedback. It seemed unwise to tell a participant "You

got some wrong" if the participant believed this was unlikely. In copying tasks, participants can self-detect about 50% of their errors (Bailey, 1978). Second, if a participant made no errors or almost none, it would not be possible to demonstrate any effects of variation in the independent variable. Much effort was expended in ten three-hour procedural exploration sessions developing a task that would produce error rates between 1 and 18 percent. The average achieved was about 9 percent. It was felt that too high a rate would be as insensitive as one too low. There was a considerable range of inter-subject differences.

The task finally selected obviously exploited the limits of short-term memory. The use of random digits in measuring "immediate memory span" (Woodworth and Schlosberg, 1954, p. 696) is hardly novel. Errors increase sharply between 8 and 9 digits (Woodworth and Schlosberg, 1954, p. 705). However, in earlier experimental work and in intelligence testing, apparently individuals started to repeat the series as soon as its oral or visual presentation ended rather than after some fixed interval, as in the present study. More recent research in human information processing has exploited strings of random digits to examine the effects of rehearsal prevention--the Brown Peterson technique (Baddeley, 1976) and of rate of presentation (summarized in Kintsch, 1969), and instead of recall has examined recognition (e.g., Shepard and Teghtsoonian, 1961) and a "missing scan" procedure where the subject must say which digit is missing in a series.

Additional Methods. Other methods explored to produce appropriate error rates included the following:

1. The presentation time of a row of digits was shortened to 3 seconds. As with Bailey (1978) and in some cases of increasing the digit presentation rate (equivalent to shortening the row presentation time), a substantial number of errors could be assured through briefer displays. However, this method was rejected because (1) presentation time was measured by stopwatch and was imprecise and (2) participants too often ommitted a row entirely or fabricated one out of thin air.

2. The delay interval was shorter (5 seconds) or longer (15 seconds). Not enough errors occurred with the shorter interval, and too many with the longer.

3. With continuous rather than row-by-row display and copying of a column, defocussing the transparency projector made the displayed digits more difficult to discriminate. Though this technique was effective in producing errors (as might be expected), it seemed likely that participants might improve greatly with practice (even without item-by-item feedback), and in any case the projector could not be adjusted to produce equal blurring in all rows. It would be better to show slides of digits that had been prepared with uniform blurring.

4. With continuous rather than row-by-row display, the typewriter was placed at a 90° angle to the display so the participant had to shift vision that amount from examining the display to seeing the keys and typing paper. (Incidentally, all participants were required to type

with a forefinger so prior practice in touch typing could be ruled out.) Because participants rehearsed the digits as they shifted their gaze and the interval between looking and typing was short, this technique failed to produce errors.

5. With continuous rather than row-by-row display, the number of digits in a row was varied. Error rates were very low with 5, 6, and even 7 digits, and in some participants, due to grouping and rehearsing, they remained low even with 8, 9, and 10 digits. Individual differences were considerable, with a biochemistry technician and a supermarket checkout clerk doing especially well.

6. With continuous rather than row-by-row display, the investigator uttered random numbers while the participant was copying the column. There was no effect.

In addition, the earlier sessions led to a number of other changes:

1. The feedback display for augmented "wrong" was changed from "You got some of them wrong. Mistakes are pretty stupid in a task like this," after two participants laughed at it and said it was bizarre.

2. Instructions were changed from "Copy the numbers as fast as you can" to "Type the numbers as fast as you can" after participants reported that instructions to "copy" meant they should not make any mistakes. "Copy" implicitly conveys stimulus control. What else is there to type except what is displayed?

3. One experimental condition, in which instructions included "Try not to make mistakes," was dropped because (1) there was insufficient time to include it in a three-hour session, (2) avoiding mistakes seemed to be implicit in a copying task even when the word "copy" was changed to "type," and (3) it involved another variable than feedback. It could be introduced in subsequent research to test the effects of instructions combined with feedback.

Results

Criteria. A set of criteria for determining whether a copying error had occurred was established prior to the first data-taking session. The criteria and the categories of error to which they apply are described in Table 1. It should be apparent that errors take diverse forms, even in copying random numbers (especially when these occur in some sequence, such as a row). Presumably the different types of errors have differing origins--an interesting domain for research.

Because the effects of any information feedback presumably would not develop until there is subsequent performance, error data were taken only from the last four blocks within a condition (including the baseline conditions).

Speed of Performance. As already noted, participants were told to type as fast as possible. That was their only opportunity for self-pacing. They could not go back and make corrections, and both display time and delay time

Table 1. Error criteria in the Copying study.

1. A typed digit that differs from the projected digit in the same row position is an (confusion) error. Example: 27644310 typed, 27544310 projected.

2. If a digit at the end of an incomplete typed row matches the digit at the end of the projected row but not the digit in the same position in the projected row, it is not an (confusion) error, whether or not it is part of a sequence. Example: 401936 typed, 40193476 projected.

3. Not typing a projected digit is an (omission) error, due perhaps to an (anticipation) error. Example: 9824336 typed, 98254336 projected.

4. A typed digit is an (intrusion, addition, or perseveration) error when it occurs before the end of the row, if it does not match the projected digit in the same row position and precedes a typed sequence that matches a projected sequence. (However, it does not count as an error if there has been an omission error earlier in the row.) Example: 7276237 typed, 72723701 projected.

5. A typed digit is an (addition or perseveration) error when it occurs at the end of the row and causes the total number of digits in the row to exceed the total in the projected row, unless it is part of a projected sequence. Example: 077352809 typed, 07735280 projected. (NOT 077395280 typed, 07735280 projected.

6. A typed sequence of two or more digits matching a projected sequence is NOT considered an (location) error when the location of the sequence in the typed row differs from the location in the projected row, except where two two-digit sequences are reversed (see below). Examples: 81127813 typed, 08278136 projected or 81162781 projected.

7. Typing two projected digits in reversed order counts as one (transposition) error and is regarded otherwise as a sequence. Example: 74465901, 74456901 projected.

8. Typing three projected digits in some kind of reversed order counts as one (transposition) error and is regarded otherwise as a sequence. Example: 366 typed, 663 projected; 068 typed, 806 projected.

9. Typing three projected digits of which two are the same so the single digit is doubled counts as two (exchange) errors and is regarded otherwise as a sequence. Examples: 363 typed, 636 projected; 336 typed, 663 projected.

10. Typing four projected digits in which two two-digit sequences are reversed counts as one (transposition) error and is regarded otherwise as a sequence. Examples: 3663 typed, 6336 projected; 3366 typed, 6633 projected; 3654 typed; 5436 projected.

Table 1 (continued)

11. Typing four digits with some reversals of the same projected digits counts as two (transposition) errors and otherwise is regarded as a sequence. Examples: 2415 typed, 1425 projected; 5142 typed, 1425 projected.

12. Typing four projected digits of which three are the same, so the single digit is tripled, counts as three (exchange) errors and is regarded otherwise as a sequence. Example: 3666 typed, 6333 projected.

The omission or repetition of a row or typing a row unrelated to the projected row does not constitute an error, but in calculating error percentage the denominator must be reduced if the row is omitted; or the error total should be reduced proportionately. A repeated row should be disregarded. If the typing clearly jumps within a typed row from one projected row to the next, the typed row will be regarded as a legitimate one composed of the two projected parts, and one row will be indicated as omitted. Non-digit characters per se are not errors.

were fixed at 5 and 10 seconds, respectively. Thus, it should not be surprising that within participants and between conditions across and within participants there was little variation in the time a participant took to copy a block of random numerals. Median durations were derived from all five blocks per participant in a condition. Across conditions the medians (of block medians) varied only within a 9-second span; the median of these was 3 minutes and 14 seconds. The median of the medians among participants was 3:12; individual medians ranged from 2:51 to 3:32, presumably reflecting differences in typing speeds or, possibly, further rehearsal during typing.

Accuracy of Performance (Errors). As Table 2 indicates, the overall mean error rates of the three consequation/feedback conditions were almost the same: 9.5 percent for "most right,", 8.9 percent for "some wrong," and 9.0 percent for "some wrong. Mistakes...most unfortunate..." By inspection, these results justify no assertion of different effects between feedback conditions and fail to warrant statistical significance treatment. The error rates come close to those in the prior session for the six participants who took part in that (rather than in procedural exploration sessions): 9.7 percent, 8.4 percent, and 9.2 percent.

The error percentages for the two baseline conditions, without feedback, were 11.2 (10.8 in the prior session) at the start of the session and 7.0 (6.8 in the prior session) at the end. The relatively small differences between these and the feedback condition percentages appeared to have little material importance and if only due to marked within-subject-condition variance were not examined for statistical significance. They may well have reflected improving performance during the session as a result of practice, including procedural improvements despite efforts to forestall this by means of prior and exploration sessions. This hypothesis is supported by the error percentages according to the order of condition encounter: 11.2, 11.0, 9.1, 7.4, and 7.0, as well as by the difference between a 6.8 baseline mean percentage at the end of the prior session and a 11.7 percentage for the same six participants in the baseline mean percentage at the start of the principal session six to 14 days later.

The low error rate in the final baseline condition, without feedback, may have been due to the cumulative effects of the preceding three feedback conditions, as well as, or instead of, being due to practice. The drop in this final percentage prevents any assertion that feedback reduced errors, since it was added as a within-subjects control condition. It could be hypothesized that it might have risen, as it did between sessions, and that differences might have been demonstrated among feedback conditions, if (1) there had been intervals longer than 5 minutes between conditions, and (2) there had been more than five blocks per condition (as in better time-series experiments).

The overall error percentage in the study, for the principal session, was 9.1. This represented success in developing ways to assure error rates that were neither excessively low nor extremely high. Error percentages varied considerably among participants, from 2.1 to 15.6. The 2.1-error participant was assigned 8-numeral instead of 9-numeral rows through investigator misinterpretation of the ground rules, as that participant had had a

Table 2. Copying Errors. Figures in parentheses show errors in prior session.

Participants	Consequation Conditions					Total Errors	Number of Numerals	Percent Error
	A	B	C	D	A			
I*	20	6	18	23	4	71	1800	3.9
II*	41	17	44	40	26	168	1800	9.3
III	55	57	54	50	33	249	1600	15.6
IV	42	47	21	33	26	169	1600	10.6
V*	76(18)	27(18)	33(18)	16(33)	33(13)	185(100)	1800(1600)	10.3(6.3)
VI	9(24)	2(29)	5(23)	11(26)	6(19)	33(121)	1600(1600)	2.1(7.6)
VII	33(58)	47(45)	10(46)	28(34)	17(23)	135(206)	1600(1600)	8.4(12.9)
VIII	13(28)	60(19)	38(24)	34(19)	34(23)	179(113)	1600(1600)	11.2(7.1)
IX	58(58)	28(56)	44(34)	38(49)	23(37)	191(234)	1600(1600)	11.9(14.6)
X	26(43)	26(23)	27(20)	25(20)	32(19)	136(125)	1600(1600)	8.5(7.8)
Totals	373(211)	317(190)	294(165)	298(181)	234(134)	1516(881)	16600(9800)	
Percents	12.2(10.8)	9.5(9.7)	8.9(8.4)	9.0(9.2)	7.0(6.8)			9.1(9.0)

*These participants copied 9-numeral instead of 8-numeral strings. Participant V copied 8-numeral strings in the prior session.

7.6 percent error rate in the prior session with 8-numeral rows. One participant who had a 6.3 percent error rate in the prior session with 8-numeral rows had a 10.3 rate with 9-numeral rows in the principal session, bearing out the conclusion from the procedural exploration sessions that error rate increased with the number of numerals in a row to copy (and to remember after a 10-second interval).

No analysis was performed on the incidence of different kinds of copying errors. Such an analysis (see Bailey, 1978, for a limited example) would become an appropriate aspect of an information processing investigation of the origins of different kinds of copying (and short-term memory) errors involving random numbers.

Self-Report Data. Each participant responded to eight queries from an investigator at the end of the participant's final session. The queries and replies were:

(1) What did you think was the purpose of the experiment today? Two participants said simply it was to determine accuracy, in clerical jobs (IX) and in remembering (IV). The others all referred in some fashion to the effects of the feedback, as "difference in accuracy with different types of evaluation" (I), "what rewards would bring best efforts" (VIII), "something to do with positive and negative reinforcement--but I tried not to let it interfere" (VII), "Effect of positive, negative reinforcement" (V), "how better one would do with a stimulus like wrong" (III), "see if any change occurred in the number right or wrong after a statement right or wrong" (II), "like something in psychology, stress; I did react--reacted more when scolded" (X), and "to see what someone in authority would say and how it would affect your work" (VI). Thus, most verbalized the aim of the study.

(2) What did you do during the delay period, between seeing the numbers and typing? Two said merely that they "tried to memorize" (VI) or "tried to remember" (IX) the numbers. One (V) "figured out ways I could remember them best." One (VIII) reported holding "a mental picture of the numbers," and two (II and VIII) tried, respectively, to visualize them or the easiest four. Including these two, six participants said they repeated the numbers (to themselves or in their heads); for (VII) these were "the hardest four." Three of these said they separated a row of numerals into groups or blocks. One participant (III) reported turning some numbers into dates, another (I) into dates, addresses, and telephone numbers, and this last reported also using "a rhythm technique and other devices, slowing the rhythm in typing, and "telling myself, 'let it go' when I made a mistake." On the basis of these self-reports, one might state that the study investigated short-term memory rather than copying, which was simply the method of inquiry.

(3) What was your reaction to the comments shown you at the end of typing each column? Only two participants reported that these led to performing differently: IX said "I tried to do better, not make the same mistakes," and IV, who encountered condition B second, said "At first I felt good, then not so good; they made me work harder, concentrate more," but II said I'm not sure I tried that much harder when it said I got some of them wrong."

Three reported a kind of disregard: X said that "at first I took it literally, later I took my mind off it so I wouldn't get depressed"; I said "my thoughts were to disregard them--with the positives I was anxious to get to the next sequence, with the negatives I told myself to ignore them--I got internally kranky about them--I was hungry and thinking about things to do at home--I wanted to finish"; III said "I thought it was kind of funny--I even snickered--but I realized it was an experiment." Five participants expressed feeling a lack of congruence between the comments and what they had done: VII said "I didn't think it always fit"; VI said "sometimes I thought they didn't say what I did"; V said "a couple of times I thought I had them all" (though V did not); II said "no matter what I typed I could have gotten the same correct"; and VIII said "I couldn't figure them out--there was no relationship to what was happening; they'd appear the same no matter what my performance was." If these self-reports reflected the effects of the feedback comments on error-making performance, little or no impact would be presumed.

(4) Consider a 7-point scale on which very unpleasant is 1, very pleasant is 7, and neither is 4. Rate how you felt about being told (a) You got most of them right; (b) You got some of them wrong; (c) You got some of them wrong. Mistakes with numbers can be most unfortunate in some clerical tasks. The results are shown in Figure 1. Clearly, feedback (a) had no or little reported hedonic effect on five participants, though the reaction from IV is consistent with replies to query 3, and feedback (b) had no or little reported hedonic effect on six participants. Feedback (c) had a stronger effect. Indeed, III said "it was threatening," but VIII reported being mystified about the reference to clerical tasks. Participant VI said (a) "was nice," and commenting on (a) and (b) participant VII reported relying more on self-judgment, "telling myself when I didn't do well." By inspection, there appears to be little relationship between these ratings and error totals for individuals in different feedback conditions, though some relationship might have been expected.

(5) Did you try at any time not to make mistakes? All participants replied affirmatively and four said all the time, though II indicated beginning to do so about one-third of the way through the session; participant I denied trying harder at one time than another but said it was harder to concentrate on particular numbers when there were negative comments. In response to Why?, three (I, V, and IX) mentioned "challenge," I added "I wanted to prove to myself I could do it all right," VIII said "fun," X said "self-satisfaction," and IX and II, respectively, "because I don't like making mistakes" and "because basically I'm being tested; I'm conditioned to do things right." These self-reports could be taken as evidence of either intrinsic motivation or past conditioning that could well have overridden any differential effects of the feedback comments. In response to How?, four participants (I, IV, VIII, and IX) mentioned "concentrating" (to which I added "paying attention only to numbers"), two (II and VI) mentioned "trying hard" to remember or memorize, II mentioned repeating the numbers, and VII reported "trying to find the best method to keep numbers in memory longer." Together with the replies to Question (2), it seems apparent that participants tried and presumably succeeded in developing or instituting procedures to improve their memory

Figure 1. Ratings of consequators. Roman numerals identify participants giving post-experiment ratings at particular positions on the scale. Participants selected numbers between 1 and 7 rather than seeing a graphic scale.

		1	2		3	4	5	6	7	
(a) You got some of them right.	Very Unpleasant					IX, VIII, VII	V, I	X, VI, III, II	IV	Very Pleasant
(b) You got some of them wrong.	Very Unpleasant		V, IV, III	I	IX	VIII, VII, VI, II				Very Pleasant
(c) You got some of them wrong. Mistakes with numbers can be most unfortunate in some clerical tasks.	Very Unpleasant	V, III	X, IX, IV	I	VI	VII, II	VIII			Very Pleasant

Neither (at 4)

(and copying) performance. These procedures could include "concentrating" and "trying hard" (terms that need further analysis). Such procedural changes can be viewed as a form of "learning" (though perhaps different from what that term is often assumed to mean). If this were the case, the design of the study failed to prevent within-study learning from confounding the effects of the independent variable (type of feedback).

(6) Did you ever type more slowly to keep from making mistakes? Only two participants, I and X, denied ever typing more slowly, though some said doing so was only occasional, and two reported they "must have" or "might have." As noted earlier, the duration data showed almost no variation in the median times taken to copy columns of numbers, between conditions. Within conditions, durations of copying individual columns rarely reached a difference of 5 percent greater or less than the median.

(7) Did you think any sets of numbers were repeated? Six participants said they thought so, but some thought there were only a few or once; one said "no," one didn't know, and a third reported thinking so but being unable to detect any repetitions. In actuality, as noted earlier, within either the principal or prior session there were no repetitions of rows of numerals (though there were chance repetitions of short sequences). However, for the most part the columns (blocks) of rows used in the prior session were used again, in different conditions, in the later one, at least six days later.

(8) Comments? Participant II expressed indifference to the negative statements, as "not disturbing." This individual reported having taken a course in introductory psychology. Participant X said "it was a very good test, it makes you think all the time." In view of the latter comment, together with replies to Question 5, it would seem that the experimental situation hardly resembled everyday clerical or working circumstances. Perhaps greater external validity--and possibly more differential effects from the kinds of feedback--would have resulted from maintaining each feedback condition for a much longer time, as in many single-subject, time-series designs.

Discussion

As noted at the outset of this report, this Copying study was exploratory, in its principal session as well as sessions that were devoted explicitly to procedural exploration. Due perhaps to the design of the experiment, the data gave no reason for saying that any consequation/feedback condition was more effective than any other for reducing copying errors, or more effective than no feedback; for the same reason, there can be no statements about ineffectiveness.

How might a consequator change error-producing behavior? According to reinforcement principles, a favorable consequator after a performance error should, by itself, strengthen that behavior (make it more likely to reoccur), an unfavorable consequator should weaken it (make it less likely to reoccur). But the consequator must be contingent on the specific performance it is supposed to influence (not a series of actions in which the erroneous response

is imbedded. The contingency may be specified through a discriminator, which identifies the erroneous action, and that discriminator may be presented to the performer (or originate internally from past experience) at the same time as or as part of the consequator (or preceding or during the performance). Such action-produced discriminators are what is known as information feedback or knowledge of results. Another kind of discriminator identifies the "correct" action to take; this, too, may originate externally (or internally), after the action (or before or during it). A different action in the future is one result, self-correction another.

None or little of this analysis applies to the consequation/feedback conditions in this Copying study. Verbal consequators were given at the end of a block of 80 or 90 numerals rather than for a numeral by itself. A consequator was not contingent on a specific action that re ulted in an error, nor were there external discriminators that identified those. (There undoubtedly were some internal self corrections but fewer than those described by Bailey (1978), since participants could not go back and change anything they typed.) Some of the participants, in response to Question 3 at the end of the principal session, expressed their concern that there seemed to be little relationship between the feedback and what they had done.

On the other hand, in accord with reinforcement principles a consequator may influence the adoption or change of the "rules" that determine what human beings do in rule-following behavior. Rules are procedures. Procedures are sequences of actions with sequences of discriminators or they may be complementary behavior that affects or modifies the performance of direct interest. In this Copying study, according to self-reports participants adopted procedures to prevent mistakes, such as rehearsal (covert repetition of numbers during the delay period), grouping (chunking) of numerals (and even rhythm), visualization, and concentration. What explicit effect did the consequation/feedback have on adopting and following these tactics? Under the circumstances, neither a verbal consequation contingency nor an external discriminator can be identified as responsible. The procedures may have come from prior experience in memorizing verbal material. More investigation of such matters is needed, with reliance on performance data as well as self-report--valuable as the latter can be for exploration.

The feedback conditions may have added to the "demand characteristics" of the situation, but the potentiation already present for participants in being, in a sense, "tested" in an experiment probably affected them in the baseline conditions as well. It is clearly difficult to study error-making performance in a laboratory situation. Effects of prior experience cannot be fully forestalled. Students especially tend of try to avoid making mistakes; their academic environment requires such behavior. Additional variables may have less observable impact than they would in other circumstances.

CONSEQUATOR RATING STUDY

It may be possible to determine how much influence a consequator would exert over some behavior of an individual by asking him or her whether or how much he/she likes or dislikes a particular outcome of that behavior, or otherwise reacts hedonically to an outcome. Presumably if an outcome is reported as liked or wanted, or as making a person happy, it will function as a favorable consequator and make the behavior more likely to reoccur. If it is reported as disliked or rejected or as making a person unhappy, it will function as an unfavorable consequator and make the behavior less likely to reoccur. The technique of asking people what they want has been used in some token economy studies to determine what tokens should be exchangeable for, and how many tokens per ultimate consequator. However, it does not seem to have been exploited to find out what could "punish" or suppress some behavior, or to discover outcomes people might try to avoid. Nor has it been investigated with respect to verbal consequators, favorable or unfavorable.

To explore reactions to such potential consequators, seven working women were asked to rate how unhappy or happy they would feel about another person's verbal comments concerning a task they had performed. The participants all worked in various offices in a technical/industrial park in a Washington, D.C. suburb, in clerical or supervisory positions. Their ages varied from the twenties to the fifties. They were individually queried after receiving written instructions. Anonymity was assured, and they signed informed consent forms.

They were shown a graphic, bimodal, 9-point (9-digit) rating scale on which the end digit 1, always on the left, meant that the comment "would have made you feel extremely unhappy," the end digit 9, on the right, meant that the comment "would have made you extremely happy," and the middle digit 5 meant that "it wouldn't have made you feel one way or the other." A participant selected and wrote down on a separate blank sheet of paper a rating number for each item, working at her own pace but being timed. There were 124 items listed alphabetically in columns on other sheets of paper. They are listed by ratings in Table 5, but their alphabetical order in the study indiscriminately mixed presumably unfavorable, neutral, and favorable comments. The items were drawn from a thesaurus and from the investigator's own head. They included adjectival statements and phrases, nouns, and exclamations, some adversely critical, some complimentary, some neither. Some were oriented primarily towardthe task itself and its outcome, some toward the individual's personality characteristics as evidenced in task performance.

There were two tasks to consider (separately): (1) performing a job at work, and (2) taking a test at school. For the first, a participant was told to "Suppose you asked an associate whose judgment you respected to comment on a job you had just done, and the associate made one of the following comments." For the second, the participant was told to "Suppose when you were a student you took a test and the teacher, using some objective criterion, made one of the following comments."

Further, each of these tasks had three conditions for each participant: (1) there was no mention whether it was easy or difficult, [illegible] it was

Table 3. Median ratings of happiness-unhappiness about an associate's comments on performance of a difficult job. (Rating of 1 means extremely unhappy, of 9 extremely happy, of 5 neither unhappy nor happy.)

Rating for Difficult Job	Comment	Rating for Easy Job	Remarks
1	Bad stuff	1	Unanimous (E)
1	It's deplorable	1	(b), (g)
1	It's disappointing	1	(g)
1	Disapproved	1	(g)
1	It's discreditable	1	(e), (g)
1	It's disgraceful	1	
1	It's dreadful	1	(b) Unanimous (E)
1	You're dumb	1	(b) Unanimous (E)
1	Dummy!	1	Redundancy, Unanimous (E)
1	It's a failure	1	
1	It's frightful	1	(b), (g)
1	It's ghastly	1	(a), (b)
1	Horrible!	1	Unanimous (E)
1	Idiot	1	Unanimous (E, D)
1	It's idiotic	1	Redundancy, Unanimous (D)
1	Imbecile!	1	Unanimous (E, D)
1	It's imbecilic	1	Redundancy, Unanimous (E, D)
1	It's incompetent	1	(g)
1	It's lousy	1	
1	Moron!	1	Unanimous (E, D)
1	It's moronic	1	Redundancy, Unanimous (E.D)
1	Poorly done	1	(g)
1	Poor stuff	1	(g), Unanimous (E)
1	It's ridiculous	1	
1	Shame on you!	1	
1	It's stupid	1	
1	You're stupid	1	Redundancy, Unanimous (E)
1	It's terrible	1	Unanimous (E)
1	It's very bad	1	Unanimous (E)
1	It's worse than usual	1	(e), (g)
1	You're wrong	1	(e), (g)
2	It's absurd	2	
2	It's awful	1	(a)
2	Below par	1	
2	You're crazy	1	
2	It's defective	1	(a)
2	It's disreputable	1	
2	You're foolish	2	
2	It's irrational	1	
2	You're nutty	2	
2	It's pitiable	1	(a), (g)
2	Pretty bad	1	
2	It's unsuccessful	5	

Table 3, Continued (2)

Rating for Difficult Job	Comment	Rating for Easy Job	Remarks
3	It's false	1	(e)
3	It's faulty	2	
3	It's imperfect	3	(e)
3	It's inaccurate	1	
3	It's inadequate	1	
3	It's incorrect	1	(e)
3	It's insufficient	1	(a)
3	It's inferior	1	(a)
3	You're obtuse	3	
3	It's second rate	2	
3	You're silly	1	(e), (h)
3	It's so so	5	(f)
3	It's unfortunate	2	
3	It's unobjectionable	4	(a)
4	It's contemptible	1	(a), (c), (d), (e)
4	It's mediocre	4	
4	You're mistaken	1	
4	No comment	4	
4	Not good	3	
4	It's reprehensible	3	
4	It's tolerable	3	
4	It's uneven	4	
5	It's accurate	7	(b), (g)
5	It's adequate	5	(a), (c)
5	It's average	5	(a), (c)
5	It's competent	5	(g)
5	It's creditable	5	(a), (c)
5	It's fair	3	(h)
5	It's fine	5	(a), (b), (e)
5	It's good enough	5	
5	It's laudable	5	(a), (b), (d)
5	It's middling	4	(f)
5	It's normal	5	
5	O.K.	5	(b)
5	It's ordinary	5	
5	It's rational	5	(a)
5	It's sound	5	
6	You're astute	5	(a), (e)
6	Better than usual	7	(a), (c), (d)
6	It's correct	5	(a)
6	It's effective	6	(a)
6	It's encouraging	7	
6	It's nice	5	(a)
6	Not bad!	5	
6	You're right	6	
6	Thanks	5	

Table 3, Continued (3)

Rating for Difficult Job	Comment	Rating for Easy Job	Remarks
7	It's distinguished	7	(b)
7	Good for you	7	
7	Pretty good	6	
7	It's true	5	(a), (h)
7	It's worthy	5	
8	Above par	5	(a), (c)
8	It's admirable	7	
8	Approved	5	(a), (c), (d)
8	It's commendable	6	
8	It's exemplary	5	(e)
8	It's first rate	7	(a)
8	Good stuff	7	
8	It's perfunctory	5	(a), (c)
8	It's praiseworthy	6	
8	You're sharp	7	
8	Well done	7	
9	Bravo!	9	
9	You're bright	7	
9	It's brilliant	9	Redundancy, (f), Unanimous (D)
9	You're brilliant	9	
9	It's excellent	9	
9	Genius!	9	(f), Unanimous (D)
9	It's grand	8	
9	Great!	9	(f)
9	It's magnificent	7	(a)
9	It's meritorious	6	(b)
9	It's perfect	9	(f)
9	It's splendid	8	
9	It's a success	6	(a), (h)
9	It's superb	9	(f), Unanimous, (D)
9	It's superior	9	Unanimous, (D)
9	It's superlative	8	
9	It's tip top	8	(h)
9	It's very good	8	
9	It's wonderful	8	

KEY to Remarks

a: At least one E rating is 4 places from the median.
b: At least one D rating is at least 4 places from the median.
c: There are extreme E ratings above and below the median.
d: There are extreme D ratings above and below the median.
e: There are 3 ratings above the E median, none below it.
f: There are 3 ratings below the E median, none above it.
g: There are 3 ratings above the D median, none below it.
h: There are 3 ratings below the D median, none above it.
E: Easy Job ratings.
D: Difficult Job ratings.

described as easy ("Now suppose you thought the job (test) was an easy one"); or (3) it was described as difficult ("Now suppose you thought the job (test) was a difficult one"). Each participant went through the list of 124 items at least six times, once for each of the three conditions in each of the two tasks. The three conditions for the Job, in the order of no mention, easy, and difficult, were followed by those for the Test, in the order of no mention, difficult, and easy. In addition, the no-mention condition for the Job was repeated with six participants, in five cases after the Test portion, in one between the Job and Test portions; and Job Difficult was repeated with one of the participants, at the end.

Results

Durations. The average duration for completing the list of 124 items was about six minutes but individuals varied greatly. The slowest respondent took about 13 minutes for the first set of reactions to the 124-item list, and about 7.5 minutes for the last; the fastest respondent took 5 and 4 minutes. Durations were reduced by 20 to 50 percent between first and last, with all participants speeding up during the session.

Reliability. The first encounter with the list of items and rating them was the condition of Job with no mention of ease or difficulty. The data were not used except to compare them with the later Job No-Mention. In that comparison, 47 percent of the 744 ratings of items for six participants differed in the later set from the earlier; 23 percent of the differences were greater than one place on the scale. There were approximately the same number of increases in ratings (25 percent) as decreases (22 percent). Among individuals the differences ranged from 30 percent (with 14 percent greater than one place) to 65 percent (with 48 percent greater than one place). When one subject repeated the Job Difficult ratings, there were 22 percent fewer differences than there had been between the two Job No-Mention conditions. It might be presumed that the first Job No-Mention condition, with which a session began, was relatively unreliable, and that reliability subsequently increased. Due to the remaining unreliability, inter-subject variance, and small number of participants, the study must be viewed as exploratory. Analysis of statistical significance was not undertaken.

Item Ratings. The median (between participants) ratings for the 124 items in the Job Difficult condition are shown in Table 3. The table also shows the ratings for the Job Easy condition. The remarks column contains a number of data modifiers or cautions. In cases of (a) and (b), where at least one E rating or D rating is 4 places from the median, and in cases of (c) and (d) where there are extreme E and D ratings above and below the median, it may be conjectured that the item presented some sort of difficulty to one or more participants, resulting in a deviant rating. In any case, one or more participants had an idiosyncratic interpretation of the item. Instances of (e), (f), (g), and (h) indicate there would be somewhat different results if means had been derived instead of medians: (e) and (g) indicate that the mean rating would be higher, (f) and (g) that it would be lower. These modifiers can be related to differences between Easy and Difficult ratings (see below). The (a), (b), (c), and (d) indicators should caution about the reliability of these items in any future use.

Some comments that might be presumed to be just information-carriers may be seen to occasion self-reports of extreme unhappiness. These include "Wrong," "Incorrect," and "False," though, as noted below, their antonyms are relatively neutral. "Failure" and "Success" both exert strong hedonic effects.

Job Easy vs. Job Difficult. Among the 124 items, there were 62 differences (50%) between medians for Job Easy and medians for Job Difficult. Among these 62, 90 percent (56) were instances where the rating was higher in Job Difficult than in Job Easy. Among these 56, 28 were ratings of comments that could be called favorable (rated above 5 in Job Difficult), 24 were ratings of unfavorable comments (below 5), and 4 were ratings of neutral comments (at 5). Thus, higher ratings for unfavorable plus neutral comments were just as many as higher ratings for favorable comments. A comment that was favorable for task outcome when the task was easy was more so than when the task is difficult; the individual reported feeling happier. A comment that was unfavorable for task outcome when the task was easy was less so when the task was difficult; the individual reported feeling less unhappy. Unfavorable comments do not become more unfavorable when a task is more difficult, according to the results of this self-report inquiry.

The above summary is borne out by inspection of the data in Table 3. It will be seen that the median item ratings for Job Difficult are generally either higher than or the same as those for Job Easy; very few are lower. Although higher ones do not appear when the Job Difficult median ratings are 1, the modifier (g) for ten of these indicates that the Job Difficult ratings should be considered in the light that these had three ratings above the median and none below it and thus the means would be higher than 1. Conversely, the five (f) modifiers for items rated 9 for Job Difficult indicate that the corresponding 9 ratings for Job Easy should be considered in the light that these had three ratings below the median and none above it and thus the means would be lower than 9.

Nevertheless, many of the unfavorable comments rated at 1 were resistant to change from the easy job to the difficult job, and on some of these the respondents were unanimous in both the Easy and Difficult ratings. Examples were "Idiot," "Imbecile," and "Moron." Perhaps these were more stable because they were personality-oriented rather than task-oriented. At the other end of the rating scale, there were several favorable consequators that were rated 9 for both easy and difficult jobs, without an (f) modifier: "Bravo!," "It's excellent," and "It's superior." But none of these was rated 9 unanimously for the Easy job. If one had been, a mean rating could not have gone higher for the Difficult job. Those showing the greatest difference were "It's meritorious" and "It's a success." Some stability was demonstrated for "Neither" (neutral) ratings. Of 15 rated at 5 for Difficult, 12 had been so rated for Easy. However, of 25 rated at 5 for Easy, 12 were rated higher for Difficult (and one lower)--another indication of the trend to rate the comments higher for Difficult than for Easy.

Test Easy vs. Test Difficult. Of 868 ratings by the seven participants, 46 percent (396) were different for these two task conditions. Of the 396, 87 percent received higher ratings in Test Difficult. All but one of the participants produced higher ratings in Test Difficult, the individual increases ranging from 79 percent to 98 percent. Though these figures are not

directly comparable with the median data for Job Easy vs. Job Difficult, the 87 percent almost matches the 90 percent for differences between medians, and the 46 percent comes close to the 50 percent of medians where there were differences. Thus, the Test ratings supported the findings for Job ratings; as task difficulty increases, the same comments are reported as making a person happier or less unhappy.

Job Difficult vs. Test Difficult. When ratings for Job Difficult were compared with those for Test Difficult, of 868 paired ratings across seven participants, 393 (45 percent) were different. Of these, 65 percent (256) were higher on the Test than on the Job, and 35 percent lower. When the actual differences in scale placement were summed for the 393 pair differences, the total was 686, and 69 percent of this total (473) came from ratings higher on the Test than on the Job. All seven participants had more ratings that were higher on the Test than on the Job; six produced the same result when placement differences were totalled, and one was even. For 46 items, there were at least two more participants rating an item higher on the Test than those rating it lower, or vice versa, and of the 46, 42 (91 percent) were instances of higher rating on the Test. Among these, 19 pertained to favorable comments (rated higher than 5), 17 to Unfavorable (lower than 5) and 6 to Neutrals (5). This latter finding indicates that higher ratings for the Test situation tended to be higher whether the comment was derogatory (unfavorable) or complimentary (favorable).

Individual differences were considerable. They ranged from 27 percent more for Test (34) to 66 percent (82) out of 124 possibilities per participant (and from 56 to 144 for summed differences in scale places). The participant with the smallest preponderance of higher ratings for Test had 53 percent more than for Job; the one with the largest margin had 89 percent more for Test. The mean extent of place differences among participants ranged from 1.2 places to 2.6.

The requirement to report Test ratings called for a more retrospective self-report than did Job ratings, and the source of the consequators (comments about the task) was a teacher rather than an associate. The differences in ratings between Job Difficult and Test Difficult may have been due to these variations. What seems more striking, however, is the way in which the results for two different tasks tended to parallel those comparing easy and difficult conditions within tasks. Possibly the Test situation was reacted to as more difficult than the Job situation.

In any case, although a detailed analysis remains to be performed on the Test ratings, they appeared to resemble those for the Job data. Of particular interest is the finding that ratings were higher on the Test than on the Job for unfavorable as well as favorable comments. Ratings did not go down for unfavorable comments and up for favorable ones, as might have been suspected in comparing two types of tasks.

Distribution of Ratings among Consequators. The number of comments receiving Job Difficult ratings from 1 through 4 exceeded the number receiving ratings from 6 through 9, 65 to 44. This disparity may have been due to selecting more unfavorable than favorable comments in the first place, and in turn that selection might have resulted from a greater proportion of adverse

terms in the English language and thus in any thesaurus. Such a disproportion is suggested by the disproportion of hedonic terms in Roget's International Thesaurus (Third Edition). Under "pleasantness" are 47 lines of nouns, verbs, adjectives, and adverbs, but under "unpleasantness," 90 lines; under "pleasure," there are 82 lines, but under "displeasure," 149. There is no section for "like," but under "dislike" there are 41 lines. In addition, there are 25 lines under "approach" but 106 under "avoidance," not to mention 47 under "escape," 40 under "abandonment," and 23 under "rejection." (However, one finds 167 lines under "desire." "Aversion" is listed under "dislike.") More inquiry seems warranted as to the distribution of favorable and unfavorable verbal consequators in spoken and written English (and other languages).

There were almost as many median ratings (50) at the extreme places on the scale--1 and 9--as in intermediate places (59), excluding the central place, 5. This distribution too, may reflect the selection process, which in turn may reflect language practice. Or it may result, at least in part, from behavior in using a 9-point rating scale, especially with regard to the small number of 4 and 7 ratings.

Antonyms. Interesting comparisons may be made among antonyms. For example, for Job Easy, "Approved" was rated 5 (i.e., making the respondent neither happy nor unhappy), whereas "Disapproved" was rated 1 (i.e., making the respondent extremely unhappy.) The same contrast occurred with "It's correct" and "It's incorrect," "It's rational" and "It's irrational," It's true" and "It's false," "It's competent" and "It's incompetent," "Above par" and "Below par," "It's creditable" and "It's discreditable," and "It's adequate" and "It's inadequate." A tendency to rate the "positive" term at or close to a neutral position while its "negative" antonym remained at the extreme end of the scale was evident, to a smaller degree, among Job Difficult ratings and among other consequators, such as "You're right" vs. "You're wrong" (6 and 1 for both Job Easy and Job Difficult), "It's accurate" vs. "It's inaccurate" (7 and 1 for Job Easy), "Well done" vs. "Poorly done" (7 and 1 for Job Easy, 8 and 1 for Job Difficult), "Pretty good" vs. "Pretty bad" (6 and 1 for Job Easy), "Better than Usual" vs. "It's worse than usual" (7 and 1 for Job Easy and 6 and 1 for Job Difficult), "It's a success" vs. It's a failure" (6 and 1 for Job Easy), and "Good stuff" vs. "Bad stuff" (7 and 1 for Job Easy and 8 and 1 for Job Difficult). Studies which have used "right" and "wrong" (or similar terms) for information feedback about performance presumably have assumed equivalence in their hedonic (and consequation) status, if they have made any such assumptions at all, but these appear to be questionable. In particular, many terms that might be assumed to be hedonically positive (and thus favorable consequators) turned out, in this study, to be hedonically neutral, or close to it, especially when respondents reacted to the task as an easy one; these terms included "It's fine," "O.K.," "It's correct," "You're right," and "Thanks."

Task vs. Personality Terms. The extreme reactions to personality terms such as "Idiot" (and also "Genius") have already been noted. The "It's" items might be regarded as task modifiers and the "You're" terms (including those so implied, such as "Moron!") as persen modifiers. The comments presented to participants failed to make it possible to compare systematically these two

types. Indeed, those that the investigator had supposed to be hedonically neutral and that were rated as such were all "It's" comments. The distinction deserves further inquiry.

Individual Differences. With a sample of only seven participants not too much can be said about differences between individuals, except that even such a small sample revealed substantial differences. Earlier summaries of data have indicated some of these. Some were probably due to initial unfamiliarity with the rating scale. The ratings for Job Easy, which was rated before Job Difficult, had 26 instances of at least one rating four places from the median among participants, whereas the Job Difficult ratings had only 8, and there were 10 instances of extreme ratings both above and below the median for the Job Easy task, 4 for Job Difficult. One participant showed a strong tendency to rate items 9 when others rated them lower. Some rated an item 9, others 5. Some terms may have created rating problems by being relatively obscure to some individuals, such as "ghastly," "laudable," "perfunctory," and "reprehensible." Occasional individual ratings contrasted markedly with those of other participants, such as a 1 for "It's magnificent," and a 9 for "It's unobjectionable." Comments that produced idiosyncratic responses or marked variance in responses should be purged from the list for any subsequent use.

Discussion

Though this rating-scale study has not demonstrated that verbal reactions by someone else to a person's task performance can function as verbal consequators to influence the individual's subsequent performance, it opens the way for investigating that hypothesis. As common experience tells us, it was shown that such verbal reactions differ in the extent to which they affect people hedonically, that is, the ways and levels in what people say they feel. The empirical question is the degree to which such verbal comments exert influence as consequators commensurate with their influence on hedonic self-reports. Self-report measurement through rating scales can provide some basis for selecting potential verbal consequators to investigate this question.

Although this exploratory study should be replicated on a wider scale, with more participants and some revisions in the verbal reactions, the present study has suggested several significant factors to consider both in a more definitive inquiry and in experimentation on the behavioral influence of verbal consequators.

(1) A comment from another person will be reported as arousing different amounts of feeling according to the stated (and perhaps experienced) difficulty level of the task toward which the comment has been directed.

(2) The same complimentary comment will produce greater reported happiness when the task was a difficult one than when it was an easy one.

(3) The same derogatory comment will produce less reported unhappiness when the task was a difficult one than when it was an easy one.

(4) The same comment can produce differences in reported happiness or unhappiness when tasks themselves differ (though this conclusion was confounded, in this study, by a difference between the sources of comments as well--a variable that merits further investigation).

(5) Some terms (e.g., "wrong") that have been employed as information feedback in studies of the effects of knowledge of results seem to exert strong hedonic effects--which may or may not have confounded the information variables in those studies, since their antonyms (e.g., "right") appear to have little or no hedonic influence. In short, antonymous verbal reactions do not necessarily exert equivalent hedonic impact, partly because the complimentary member of the pair may produce little of this whereas the derogatory member may exert a great deal. How general this bias is remains to be ascertained.

(6) People differ considerably in their rating reactions to the verbal reactions of others about their task performance, even within a small group of working women. The causes of such differences presumably extend beyond differences in handling a rating scale and in familiarity with terms.

Rating scales of potential verbal consequators might serve another purpose besides the experimental investigation of consequation variables. They might be able to distinguish among individuals with respect to their susceptibility to such variables, that is, the extent to which such variables influence performance differentially among individuals. Thus, they could constitute a personality test for measuring a significant human attribute that has only recently begun to get attention among investigators of human differences, as in studies of internal-external locus of control. Ratings of verbal consequators should be related, through studies with large samples of participants, to demographic and other personality variables.

Because individuals can apply verbal consequators to themselves, and probably often do so covertly, rating scales such as those in this study might also be used to examine self-worth, self-esteem, and self-concept as these vary in task performance. Likewise it is possible they could be adapted to personnel selection in organizations.

MAPPING OF CONSEQUATION VARIABLES

One of the important by-products of experimental research is the cogitation that results from hands-on work in addition to arm-chair verbalization and along with the intellectual exploration that accompanies reading the research literature. Although consequation has often been conceptualized as a fairly simple construct (Law of Effect, Reinforcement, or Knowledge of Results), anyone starting to do experimental research in this field rapidly discovers it covers a very large number of variables and states of variables. The present investigator discovered also that the domain had never been fully mapped--that is, all the related variables and their states had never been listed in one publication, though Meister (1976) and Pritchard and Mantagno (1978) have gone further than most.

It seems important to have such a map so one will know where one is going and not get lost--that is, know what variables might be considered for manipulation and what ought to be controlled. In the course of the present project all probable consequation variables and their states were collected from a large variety of sources, carefully categorized, and published in a technical report, Variables in human consequation/feedback. This report consisted principally of four comprehensive tables. Since the technical report is readily available, its contents will not be reproduced here except to list the major headings.

Table 1: Consequation/Feedback
- A. Extent, Amount (7 subdivisions, 17 components)
- B. Comparison (4 subdivisions, 15 components)
- C. Type (2 subdivisions, 9 components)
- D. Source (5 subdivisions, 11 components)

Table 2: Action Relationships
- A. Purposes, Effects of Consequation/Feedback (4 subdivisions, 19 components)
- B. Action Aspects (6 subdivisions, 37 components)
- C. Action Relations (10 subdivisions, 35 components)

Table 3: Potentiation Relationships
- A. Potentiation Aspects (7 subdivisions, 22 components)
- B. Potentiation Relations (3 subdivisions, 11 components)

Table 4: Consequation Context
- A. Referent (3 subdivisions, 6 components)
- B. Receiver (2 subdivisions)

Note: Many components have sub-components.

A further task would be to depict, to the extent research provides evidence, how the many variables and their states are interrelated. That would be a major project in itself.

DETECTION STUDY

This exploratory study investigated how various aspects of consequation/information beedback affected performance in a detection task. The task consisted of finding on one sheet of random numbers differences (changes) from the numbers on another sheet. The numbers were printed in ten rows on a sheet. Feedback was varied in two ways: (1) An indication that the subject failed to detect a difference within a string of numerals, shown immediately to the subject reading that string. The subject received or did not receive such feedback. (2) A comment at the end of the row that appeared if the subject had not detected a difference. There was (a) no comment, (b) the comment "Miss," or the comment "Idiot." The word "Miss" was presumed to be more or less neutral hedonically, whereas "Idiot" had been rated by a number of other subjects earlier as one that would have made them extremely unhappy if an associate had used it to characterize a job they had done.

Procedures

Indications that a subject had failed to detect a difference consisted of filled rectangles that appeared under a numeral string if the subject ran a special marker device under the string. The subject was instructed beforehand to refrain from running the marker under a string in which he/she saw a difference. Thus, the filled rectangle would appear only if the subject failed to detect a difference, and the subject would get immediate feedback to that effect. There was no feedback for false positives, where the subject refrained from underlining a string where there was no difference, but these could be counted as well as the misses.

The subject was instructed that if a rectangle did appear, he/she should continue moving the marker from the end of the row to the edge of the paper. In the condition of no comment feedback, no word appeared. In the condition of comment feedback, either "Miss" or "Idiot" appeared, depending on the experimental condition. Neither of these, however, would appear unless the subject had failed to make a detection.

The numeral sheets on which the subject used the marker were premarked and preprinted with a chemical (A.B. Dick latent image process) that made the rectangles and words invisible unless the marker passed over them.

Subjects were told initially that the experiment was essentially an investigation of pattern perception. The right-hand sheet that the subject marked had different spacings between strings of numerals from the left-hand page with which the subject compared it. As the subject was informed, the spacings differed in the number of spaces and by presence or absence of a space. In addition, again as the subject was told, the order of two strings might be reversed. The reason for the instruction about pattern perception was to induce the subject to scan a line fairly rapidly as though reading a line of print, and thus to minimize the importance of detecting differences in the numerals themselves. An invisible filled circle was premarked above every location where there was a pattern difference between the [illegible] and the comparison (left-hand) page. The subject was told to run the marker above every location where he/she noticed a pattern difference [illegible] order. Such detections would then be confirmed by the appearance of the [illegible] circle. Each session began with pattern differences only.

Subsequent to ten one-hour preliminary sessions involving ten participants (see below), two participants took part in this study, and a third participation was planned. For each participant, there were nine one-hour sessions, in one case all on the same day with approximately ten minutes between sessions, in the other case with intervals between sessions ranging from a half day to several days. The data had not been processed at the time of writing this report. There were six experimental conditions, in the order ABCDEFEDC, as described in Table 1, with participants encountering seven to nine paired pages of random numerals per session. They were timed (and told so) "to see how fast you can go through each page," starting each page when the investigator said "Go" and announcing "Finished" upon completion, then proceeding to the next. They were also instructed to refrain from going backward within a row or to an earlier row. They were not told about any pattern or numeral item feedback until just before the session in which that feedback occurred, and they were not told in advance about the end-of-row word feedback. At the end of the session they replied to six queries about the experiment, including their reactions to the different kinds of feedback, which they rated on a bimodal 9-point scale extending from a statement that the feedback made the participant "extremely unhappy," through neutral, to "extremely happy."

Materials

Each sheet of random numerals had ten rows, each row averaging ten strings varying in length from four to 15 numerals in a string, with each length represented equally in frequency. Thus, each row had an average of 90 numerals, varying between 85 and 95, and an average of 12 spaces in total between strings (including double spaces). For both the original (unchanged) sheets and the difference (changed) sheets, there were 21 different selections of random numerals on the sheets, four different arrangements of strings of these, and three different orders of the rows of the foregoing. Two of the random numeral selections were allocated to each string arrangement. The differences in the orders of rows were produced by a word processor from the typed copies of the other combinations. For the "difference" sheets, there were three different locations of differences or faults. There were in total 59 difference sheets. No subject [illegible] encountered the same difference sheet twice. Each subject encountered the same order of original and difference sheets. The order was randomized.

Each of the three difference location arrangements, equivalent with respect to number and types of changes, was allocated to strings [illegible] randomly, by rotating through the list of [illegible] and [illegible] sequentially numbered strings, [illegible] a random [illegible] different kinds of changes [illegible] different [illegible] transposed numerals (transposition or reversal), [illegible] string (intrusion), extra numeral at the end of a string, [illegible] numeral within a string [illegible] space (omission), and [illegible] numeral [illegible] were four instances of each [illegible] were six [illegible] per page.

There were [illegible] types of [illegible] sals, two per [illegible] stance on each sheet of pattern [illegible]

Table 4. Detection study: Sessions and conditions.

Session 1: 9 pages. Condition A: pattern recognition only and no feedback. Combinations a through [illegible] in that order. No feedback on Changed pages. Basic instructions.

Session 2: 9 pages. Condition B: pattern recognition only with item feedback. Combinations a through [illegible] in that order. Feedback on Changed pages.* Additional instructions (1).

Session 3: 8 pages. Condition C: pattern recognition with item feedback, and change detection with no feedback. Combinations [illegible], 1 through [illegible], and [illegible]. Feedback on Changed pages.* Additional instructions (2).

Session 4: 8 pages. Condition D: pattern recognition with item feedback, and change detection with item feedback only. Combinations [illegible], 7 through [illegible], and [illegible]. Feedback on Changed pages.* Additional instructions [illegible].

Session 5: 8 pages. Condition E: pattern recognition with item feedback, and change detection with item feedback and with row feedback ("MISS"). Combinations [illegible], 13 through [illegible], and [illegible]. Feedback on Changed pages.* Additional instructions (4).

Session 6: 8 pages. Condition F: pattern recognition with item feedback, and change detection with item feedback and with row feedback ("IDIOT"). Feedback on Changed pages. Combinations [illegible], [illegible] through 24, and d. Feedback on Changed pages.* Additional instructions (4).

Session 7: 7 pages. Condition F as in Session 6. Combinations [illegible], [illegible], and c. Additional instructions (4).

Session 8: 7 pages. Condition [illegible] as in Session 4. Combinations [illegible], [illegible], and b. Additional instructions [illegible].

Session 9: 7 pages. Condition [illegible] as in Session 3. Combinations [illegible], [illegible], and a. Additional instructions (2).

*Note. There is no fault item or row feedback on [illegible] in session 2 or on the first and last pages of [illegible].

and three of one space between strings instead of two. (In the unchanged sheets, two-thirds of the spacings between strings were single spaces, one-third were double spaces). There were two instances on each sheet of the following: leaving a space and creating two strings; insertion of a space within a string, creating two strings; omission of spacing between strings, creating one string instead of two; and inserting a numeral between strings, creating one string instead of two. Pattern differences were allocated randomly in the same fashion as numeral differences. There were 14 per page.

The production process consisted of marking off rows in a table of random numbers into numeral strings of the various lengths. String lengths were allocated by rotating the ten string lengths (four to 13 numerals) through the approximately 100 string positions on a sheet with a random number table. As noted, there were four different string arrangements, each with two random numeral selections. The eight divided-up random number tables were typed into a word processor. The numeral and pattern differences were then put into the word processor, after it first produced 10-row original sheets without them, to create the difference sheets. The word processor then produced two additional row arrangements (complete vertical reversals and reversals within pairs of rows) for both the original and difference conditions. Thanks to the word processor, retyping of entire pages was not required for either the two row conditions involving reversals or for the insertion of pattern and numeral differences.

Preliminary Sessions

Due to the novelty of the technique of detecting differences among numbers and patterns in comparing two arrays of random numbers, it was necessary to conduct some one-hour preliminary sessions with ten participants (working women in a suburban technical/industrial park), to gain familiarity with the technique. Uncertainties included: (1) what proportions of misses (failed detections) might be expected; (2) how varying densities of differences (percentages of the numbers in an array) might affect the miss rate; (3) whether some kinds of differences would result in more or fewer misses than others, or perhaps virtually none; (4) how long it would take participants to compare two arrays; and (5) what would happen if time limits were placed on the comparisons.

Each participant encountered paired pages of 15 rows of random numbers each, each row containing 8 to 10 strings of 3 to 12 numerals; the string lengths were randomized on a page and there were equal numbers of each length. Among eight of the participants, each had to make a pencil mark where a numeral or spacing on one page differed from the corresponding numeral or spacing on the other page. There were 10, 15, 20, 30, or 35 differences per page. These differences consisted of ten different numeral and [illegible] changes, distributed among the differences per page in varying proportions for the different totals of page differences. Participants were told to [illegible] as they could, they [illegible] for the 15 difference [illegible] densities, they were told in advance that [illegible] to complete each row, [illegible] procedure [illegible] that they would receive [illegible] which varied [illegible]

The results showed an overall median percentage of misses (with no time limit) at 30, indicating that this technique could produce a substantial number of misses; the medians across participants for the three densities were 20 percent, 40 percent, and 29 percent. The median among participants (with no time limits) was 31, with a range from 16 to 40 percent; participant/density percentages ranged from 10 to 63 percent. There were indications that densities of 20 and 45 resulted in somewhat higher miss percentages than the density of 10 differences per page, but there was no evidence that percentages increased with density above 20. The imposition of a time limit seemed to make little difference in the total of misses or of false positives, when incompletions were not counted as misses. Time limits that gave 30 percent less time to complete a page resulted in many incompletions, and the same could occur with smaller reductions. It was apparent that imposing a time limit in this task could be a tricky operation if it was desirable to avoid incompletions, because judging the appropriate limit would be difficult. Misses were found to be distributed among types of differences within a range of 19 to 23 percent (except for two of the spacing differences, which were missed much more frequently). However, there was some variation among participants. All in all, though these outcomes were based simply on inspection of the data and should not be viewed as conclusive, they aided the investigator in developing the procedures for the study, described earlier, that would follow. As in all pilot work, they also produced some cautions: (1) Don't let participants point at changed strings with a pencil or a finger in a number-by-number scrutiny, and (2) don't select as participants accounting or purchasing personnel who work with numbers all day long, every working day.

<u>Background</u>

In human information processing, a frequent and important task is the detection of mistakes in alphanumeric materials. This task occurs in the self-detection of material being typed by secretaries and stenographers in transcribing from shorthand, longhand, or other typed copy, or being typed by the authors of the material themselves; and it occurs in post-typing proofreading by clerical personnel, authors, printers, and professional proofreaders for publishers. In naval and other military operations, failure to detect errors committed in preparing coded or textual material could result in some misfortune. Bourne (1977) found that index-term misspellings ranged as high as 23 percent in a set of computer data bases.

For this study, it was decided to take proofreading as a model of detection. There have been precedents. In their research on the effects of noise, Glass and Singer (1972) required participants to detect typographical errors these investigators had inserted into seven typed pages transcribed from *The death and life of great American cities* by Jane Jacobs. Weinstein (1974, 1977) also used proofreading errors [illegible] to test noise effects, and participants in another noise study (Moran & Loeb, 1977) had to detect word discrepancies between an oral presentation and printed text. In a more abstract fashion, [illegible] and [illegible] (19[illegible]) used the Posner-Boies technique of matching upper and lower case letters, and [illegible] and [illegible] (19[illegible]) had their experimental participants match rows of o's and c's [illegible] to detect c's.

Only two experimental investigations of proofreading [illegible] have come to the present investigator's attention and the [illegible] literature.

very sparse; an inquiry that included the local chapter of the International Typographical Union and several professional editors for organizations, as well as the Government Printing Office, and USDA Graduate School (both of which give courses in proofreading), turned up only a rather superficial text by Lasky (1954). Thompson et al. (1979) reported a study pertinent to the present one in which a typist proofread her own copy for errors and recorded the errors she found and whether the letters in which errors occurred were returned to her for retyping. This post-typing self-monitoring resulted in a considerable decrease of omission and substitution (but not grammatical) errors and of pages returned to be retyped.

The most significant psychological study found in the literature was by Crosland (1924), which was discovered after the preliminary work in this study was completed. Crosland inserted into textual material from several types of periodicals 14 kinds of typographical errors, including misspellings, punctuation errors, word omissions, and wrong type faces, with a density of about 100 or 50 per page. His 30 participants, in five groups of differing experience with printed materials, read and marked the errors they found. The overall percentage of misses for three different kinds of materials with the higher density were 23.7, 32.7, and 35.6, and the percentage for the lower density was 43.4. Individual differences were extensive. The investigator's voluminous report contains analyses of the effects of different kinds of typographical errors, different kinds of readers, rereading, and practice.

The present study was based on random numbers instead of textual material because it had a within-subject experimental design and it was necessary to obviate the learning effects that would presumably occur if participants encountered the same text repeatedly, with feedback about misses. It also seemed desirable to exclude typographical errors that might depend on an individual's educational level and that might be subject to "learning" during the study, such as spelling and grammatical errors due to ignorance. Certain other types, such as type font and type erraticity, were also excluded as infeasible with typed copy. Emphasis was placed on errors that a participant-reader would readily acknowledge as errors if they were pointed out to her--because a literate reader would assuredly have learned the correct version early in life. To get a handle on these, various sources were examined, including the errors Glass and Singer had introduced into the Jane Jacobs text. Oddly, a valuable source was the 27th annual report of the Ergonomics Research Society (1976), which contained 100 typographical errors.

Analyses of the sources showed that the types of numeral errors mentioned earlier in the description of materials in the present study represented the most numerous letter errors to be found in textual material. For example, omission or insertion of a numeral in a string of numerals could represent the omission or insertion of a letter in a word, transposition of two numerals could represent transposition of letters, and repetition of a numeral repetition of a letter. Pattern differences in the arrays of numbers were developed as analogs of textual pattern differences involving spacing, punctuation, and word reversals. Participants were asked to detect only pattern differences initially because readers of text tend to read words as units rather than the component letters; that is presumably a major reason for failing to detect component errors. It was hoped that study participants [illegible] to overlook numeral errors if their attention was [illegible] the patterns of numerals rather than individual numerals.

CODING STUDY

The purpose of this study was to find out how various kinds of verbally aversive feedback (consequation) would affect accuracy performance in a coding task. That task consisted of transformations between two alphanumeric codes, including both encoding and decoding, without elaboration or reduction and independent of categorization. The codes consisted of randomly selected numbers and letters, a tactic tnat has been favored before (Anderson, 1980). Since such codes are already well-learned in literate participants, it was assumed that feedback would affect performance only and not also involve learning. As in other studies in this project, the random selection of different but equivalent code components would make it possible to generate input materials that the same participant could repeatedly encounter without memorizing particular inputs that would make subsequent performance easier.

But as in the other studies, it had not been foreseen how difficult it would be to generate and present information processing problems to participants so they would continue to make enough errors to reflect differential effects from the states of an independent variable involving consequation. As a result, all nine three-hour sessions that were conducted, with eight participants, were devoted to trying to find a task that would occasion enough errors, and the study never reached the point of introducing the consequation variable. However, the process of seeking a viable method to assure participant errors has some interest in itself.

The process took place in two parts. In each, the codes were the eight numbers 2-9 and 24 letters (A and Z being omitted), the same eight letters occurring in each problem. In the first, with seven sessions and as many participants, a participant encountered four pairs of rows, one row of numerals and one row of letters, alternating in vertical order. The participant started with one of the numbers in the top row, identified the letter below it, found this letter in another row of eight letters, selected the numeral in the row under it, found that number in another row of eight numerals, selected the letter under it, found that letter in another row of letters, and selected the numeral under it. That was the correct answer. To register it, the participant found and marked it in an 8 x 8 matrix of numerals on the answer sheet. The participant did this for all eight top-row numerals. In these transformations (encoding and decoding), the participant relied not on memory of intercode translations but on similar positions in a row, which kept changing. Thus, the alphanumeric transformation involved position coding, that is, sequential positions. Very few errors resulted. Speeding up the process by giving small amounts of money as positive reinforcers for going faster or by taking money away for not going fast enough still failed to produce many errors. Participants reported they were trying to be accurate, though they had received no instructions to this effect.

A number of alternative schemes were tried out. (1) There were transformations from numeral to numeral and from letter to letter. That is, in the former case a participant would select a number below a number, find the corresponding number in another row, select the number below that, and so on. (2) Rows of numerals and letters were displaced so corresponding positions

in rows (the coding transformation) were not vertically in line. (3) Letters and numerals were arranged in columns rather than rows. (4) New pairs of rows were placed on successive pages of a booklet rather than on the same sheet. (5) The participant had to make transformations between three-number units and three-letter units. (6) Participants had to proceed from right to left instead of left to right in selecting numerals initially. (7) A clock code replaced the 24-letter code. None of these various arrangements raised the error rate from 0-5%.

(8) An arithmetic operation was introduced. Each time the participant made a transformation, he/she had to add or subtract one number or one letter to get the next higher or lower one in sequence for the next transformation. (9) The total number of transformations was doubled or tripled by requiring the participant to recycle through the material for each of the eight starting numerals in each problem. (10) Double recycling in combination with the arithmetic operation generated the largest proportion of errors but these included errors due to the participant's forgetting where he/she was in the recycling process, a type of error different from simply making a transformation error. In addition, this procedure required five minutes per problem of eight trials. (11) A preferable combination was a single recycle with the arithmetic operation and some monetary reinforcement to go fast. This produced 5-15% error. However, each problem of eight trials required three minutes.

It was estimated that instead of the total time originally set aside for testing the feedback variable, at least twice as much time would be needed per participant (and for the coding task) with either a larger number of shorter problems/trials yielding the lower error rates or with about the same number of problems/trials with each taking longer and yielding the higher error rates. An alternative would be to reduce the number of participants, running each of four or five (instead of ten) in a time-series multiple baseline design divided between the two arrangements.

In the two sessions of the second part of the exploration process, the single participant encountered 64 pairs of numerals and letters in blocks of eight rows and eight columns. In a random fashion, in half the pairs the numeral was first, in the other half the letter. No numeral or letter was repeated within a row. It was thought this arrangement might lead to more processing errors. The participant followed these instructions; the investigator wrote down the participant's responses:

> When I say "Go," call out the first numeral in the top row. Immediately next to it is a letter. Find that letter in the second row. Immediately next to it is a numeral. Find that numeral in the third row. Immediately next to it is a letter. Find that letter in the next row. Proceed in this way until you come to the bottom row. Call out the numeral or letter you end up with. That letter, or the letter next to the numeral you called out, is the one that corresponds to the numeral you began with.
>
> Go through this procedure as fast as you can. As soon as you have called out the bottom numeral or letter, start the procedure again with the second numeral in the top row. Repeat until you have started with all eight numerals in the top row. If you start with

the wrong numeral, I'll correct you. You will be timed for each time you go through the rows. Once you've proceeded from one row to the next, don't go back, even if you think you got mixed up.

The participant, who was not permitted to write anything down, traversed each block eight times, and completed eight blocks (64 responses) without making a mistake, except in the first two; the time to complete a block dropped by 25 percent. The participant also made transformations starting with a letter in the top row and ending with a numeral in the bottom row, with only one error in four blocks (32 responses). In one numeral-first and one letter-first block there was a space between numeral and letter within a pair and two spaces instead of one between pairs in a row and between rows. This spacing had no apparent effect on error rate or time.

In a subsequent session the task was made more complex. The procedure was similar, but the participant had to start with the two numerals in the first two numeral-letter pairs, find in the second row the two corresponding letters, paired there with different numerals, find those numerals in the next row, paired there with different letters, and so on to the bottom row of the block. Numeral-letter order still varied within pairs. With this task the participant began to produce a few errors and took almost three times longer to traverse a block. The increased load on short-term memory, combined with the different numeral-letter orders, presumably was the cause.

To test the limits, the participant was instructed to follow the same procedure but with three numerals at a time, then three letters, then three numerals, and so forth. The participant made more errors, got "lost" twice and took twice as long as before.

In a further variation of procedure, the participant went back to single numerals and letters at a time within a block but had to work with two blocks at a time. Parallel transformations had to be made in each block (and, as in all the other transformations, held in memory storage) before proceeding to the next. In one traversal the subject got lost three times and made six errors otherwise (out of 16 responses), taking about as long as in the three-numeral, three-letter procedure.

A final procedure consisted of combining the two-numeral, two-letter requirement with the two-block procedure. The participant tried one traversal, got lost, and said "I don't think I can do this."

Although this account can provide no conclusions, being based on one participant and a small number of trials, it may point the way toward a coding task with numeral and letter codes that would be difficult enough to generate errors for a test of consequation effects.

REFERENCES

Anderson, N.S. Coding strategies in short-term memory. In F.L. Denmark (Ed.), Psychology, the leading edge. Annals of the New York Academy of Sciences, 1980, 340, 1-8.

Annett, J. Feedback and human behavior. Baltimore: Penguin, 1969.

Baddeley, A.D. The psychology of memory. New York: Basic Books, 1976.

Bailey, R.W. The effects of stimulus duration and error-stage on the ability to self-detect errors. Ph.D. dissertation, Rice University, 1978.

Beatty, J. and Wagoner, B.L. Pupillometric signs of brain activation vary with level of cognitive processing. Science, 1978, 199, 1216-1218.

Bourne, C.P. Frequency and impact of spelling errors in bibliographic data bases. Information Processing Management, 1977, 13(1), 1-12.

Crosland, H.R. An investigation of proof-reader's illusions. University of Oregon Publications, 1924, 2, pp. 168-283.

Ergonomics Research Society. The twenty-seventh annual report. April, 1976.

Glass, D.C. and Singer, J.E. Urban stress. Experiments on noise and social stressors. New York: Academic Press, 1972.

Kintsch, W. Learning, memory, and conceptual processes. New York: Wiley, 1969.

Lasky, J. Proofreading and copy-preparation. New York: Mentor Press, 1954.

Meister, D. Behavioral foundations of system development. New York: Wiley, 1976.

Moran, S.L.V. and Loeb, M. Annoyance and behavioral aftereffects following interfering and noninterfering aircraft noise. Journal of Applied Psychology, 1977, 62, 719-726.

Parsons, H.M. What happended at Hawthorne? Science, 1974, 183, 922-932.

Parsons, H.M. Productivity enhancement: Using performance measurement data. Paper at the Symposium on Productivity Enhancement: Personnel Performance Assessment in Navy Systems of the Navy Personnel Research and Development Center, San Diego, CA, October, 1977.

Parsons, H.M. What caused the Hawthorne effect? A scientific detective story. Administration and Society, 1978, 10, 259-283.

Parsons, H.M. Variables in human consequation/feedback. Technical Report No. 1, ONR Contract N00014-77-C-0024. Silver Spring, MD: Institute for Behavioral Research, 1979.

Pritchard, R.D. & Montagno, R.V. Effects of specific vs. nonspecific and absolute vs. comparative feedback on performance and satisfaction. Report AFHRL-TR-78-12. Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Air Force Systems Command, 1978.

Shepard, R.N. and Teghtsoonian, M. Retention of information under conditions approaching a steady state. Journal of Experimental Psychology, 1961, 62.

Thompson, C., Lloyd, M., Bannister, B., Bonem, M., and Lack, P. The effects of self-recording, instructions and contingency contracting on typing accuracy. Paper at the annual meeting of the Association for Behavior Analysis, 1979.

Wei, W-K. and Konz, S. The effect of lighting and low power magnification on inspection performance. Proceedings of the Human Factors Society 22nd Annual Meeting. Santa Monica, CA: Human Factors Society, 1978.

Weinstein, N.D. Effects of noise on intellectual performance. Journal of Applied Psychology, 1974, 59, 548-554.

Weinstein, N.D. Noise and intellectual performance: A confirmation and extension. Journal of Applied Psychology, 1977, 62, 104-107.

Woodworth, R.S. and Schlosberg, H. Experimental psychology (revised edition). New York: Henry Holt, 1954.

END
9-87
DTIC